数理解析研究所講究録
434巻 1981年 144-153

時系列モデルと物理系における変数縮約

岐阜大 工 岸田邦治

§1. はじめに

状態変数の縮約は、たとえばBrown運動を記述するLangevin方程式の一般化としての非マルコフ的記述[1]にあるように、統計物理の基本的な見方の一つである。扱う体系としては、化学工学、原子力工学等の大規模プラントシステムや生物系などを念頭に置いているが、これら巨視系のシステム同定に最近時系列解析の手法が積極的に用いられている。その有用性については、以下のような具体例を考えれば、明らかであろう。走っている車内に振動の測定器を置いたとして、その針の動きは予測することができるであろうか。車自体は、人工物であるのでほぼその従がう方程式は既知としてもよいが、外界とつながった開放系であるために、むしろ、路面も含めて対象を暗箱として扱い、揺れの時系列データから車の振動を記述する時系列モデルを同定することでも

って、その動的性質を把握することができれば、工学として十分であろう。

ところで、この方針をさらに一歩進めて、時系列データからの情報を運転されているプラントの診断のために役立てようとすれば、時系列データから情報量規準[2]により決められた実験式(時系列モデル)を単に暗箱として扱うだけでなく、その中身も記述する動的システムの物理的把握が重要となってくる。 このような観点から巨視系における変数の縮約を考える。

§2. 時系列モデル

マルコフ過程のもとに、非平衡統計力学の *system size* 展開法[3]を用いて、さらに *normal scaling relation* の採用と定常性を仮定しうる巨視系は、線形の *Langevin* 方程式で記述できる。[4] 定常状態で運転されている工学的プラントは、このような状況にあるとみなしうることが多い。 ところで、巨視系をマルコフ過程として記述しようとすれば、十分な数の状態変数 x を取扱わねばならない。 実際の系で観測されている変数 y の数は、そのうちの一部分であることが通常である。 これを簡単に次式で表現する。

(1) $y = Hx, \qquad H = (I \;\; 0)$

I は $\ell \times \ell$ の単位行列

$\ell\,[= \dim y] < d\,[= \dim x]$

この変数の縮約のために、観測可能な状態変数の従がう方程式は非マルコフ性を帯びることになる。 さらにオンライン的な計算処理上、時系列データは時間について離散化されるために、変数の従がう方程式は、時間について粗視化される。結局、観測データ $Y(n) = (y^T(n), y^T(n-1), y^T(n-2), \cdots)^T$ が与えられた時、このデータから構成される時系列モデルと物理的基礎式との対応を調べることになる。 言い換えると、物理的に素性のはっきりとした遷移確率から定められるべき量 regression行列 K と diffusion行列 D とで表現できる理論式[5]

(2)
$$\begin{cases} x(n) = \Phi x(n-1) + f(n) \\ y(n) = H x(n) \end{cases}$$

f: Gaussian white noise

$\langle x(n) f^T(m) \rangle = 0 \quad \text{for } n < m$

$\langle f(n) f^T(m) \rangle = V \delta_{nm}$

$\Phi = \exp(K \Delta t)\,, \quad V = \int_0^{\Delta t} e^{Ks} D\, e^{K^T s} ds$

Δt: サンプリング時間

$\langle \cdots \rangle$: ensemble average

から、どのような時系列モデルが構成されるのかという問題を扱うことになる。

（2）が可観測性をもつ場合には、システム理論における Luenberger 正準形[6]を用いて以下のように（2）から $x(n)$ を消去することができる。まず、次の行列

$$N=(H^T,\ \Phi H^T,\ \cdots\ \Phi^{d-1}H^T)$$

から、d 個の独立したベクトルを選んで変換行列 T を作る。

$$T=(h_1,\ \Phi^T h_1,\ \cdots\ (\Phi^T)^{\sigma_1-1}h_1,\ h_2,\ \cdots,\ (\Phi^T)^{\sigma_2-1}h_2,\ \cdots,\ h_q,\ \cdots,(\Phi^T)^{\sigma_q-1}h_q)^T$$

ここで、σ_i は Kronecker 指数で $\sum_{i=1}^{q}\sigma_i = d$ であり、h_i は H^T の i 番目の列ベクトルである。次に、$x'(n)=Tx(n)$ なる状態空間では、（2）は次のようになる。

$$(3)\qquad \begin{cases} x'(n)=\Phi' x'(n-1)+Tf(n) \\ y(n)=H'x'(n) \end{cases}$$

ここで、

$$\Phi' = T\Phi T^{-1} = \begin{pmatrix} \Phi'_{11} & \cdots & \Phi'_{1q} \\ \vdots & & \vdots \\ \Phi'_{q1} & \cdots & \Phi'_{qq} \end{pmatrix}$$

各行列成分自体が次の行列であって、

$$\Phi_{ii} = \begin{array}{c} \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ a_{ii,\sigma_i} & \cdots & \cdots & \cdots & a_{ii,1} \end{pmatrix} \sigma_i\text{行} \\ \sigma_i\text{列} \end{array} \qquad \underset{(i\neq j)}{\Phi_{ij}} = \begin{array}{c} \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \\ a_{ij,\sigma_j} & \cdots & a_{ij,1} \end{pmatrix} \sigma_i\text{行} \\ \sigma_j\text{列} \end{array}$$

$$a_{ij,l} = h_i^T \Phi^{\sigma_i} S_{k_j - l + 1} \qquad \left(k_j = \sum_{i=1}^{j} \sigma_i \right)$$

S_i は T^{-1} の i 番目の列ベクトルである。

さらに、

$$
\begin{array}{c}
H' = HT^{-1} = \left(\begin{array}{cccc|cccc|c|cccc}
1 & 0 & \cdots & 0 & 0 & & \cdots & 0 & & 0 & & \cdots & 0 \\
0 & & & 0 & 1 & 0 & \cdots & 0 & \cdots\cdots & \vdots & & & \vdots \\
\vdots & & & \vdots & \vdots & & & \vdots & & 0 & & \cdots & 0 \\
0 & & \cdots & 0 & 0 & & \cdots & 0 & & 1 & 0 & \cdots & 0
\end{array}\right) q\text{行} \\
\qquad\qquad \sigma_1\text{列} \qquad\qquad \sigma_2\text{列} \qquad\qquad\qquad\qquad \sigma_q\text{列}
\end{array}
$$

である。そこで、$z\,x(n) = x(n+1)$ なる時間推進演算子 z を導入して、(3) から観測可能な変数と結びつかない変数を消去すれば、

$$z^{\sigma_i} y_i(n) - \sum_{j=1}^{q} \sum_{l=1}^{\sigma_j} a_{ij,l}\, z^{\sigma_j - l} y_j(n)$$

$$= \sum_{l=1}^{\sigma_i} h_i^T \Phi^{l-1} z^{\sigma_i - l + 1} f(n) - \sum_{j=1}^{q} \sum_{k=1}^{\sigma_j - 1} \sum_{l=0}^{k-1} a_{ij,\sigma_j - k}\, h_i^T \Phi^{l} z^{k-l} f(n),$$

となる。最終的に $M = \max(\sigma_i)$ とおけば、上式を形式的に次のように書くことができる。つまり、

(4) $$A(z)\, y(n) = B(z)\, f(n)$$

$$A(z) = \sum_{k=0}^{M} A_k z^{M-k}, \qquad B(z) = \sum_{k=0}^{M-1} B_k z^{M-k}.$$

z の定義から (4) は、次のようになる。

(5) $$\sum_{i=0}^{M} A_i\, y(n-i) = \sum_{i=0}^{M-1} B_i\, f(n-i).$$

これは表現が異なるが、減衰理論でいう非マルコフ的方程式である。しかし、右辺の f は観測データから構成できない量である。そこで、観測データのもとに閉じた式とするた

めに、fと等価な雑音源たる次のinnovationを導入する。

(6) $\quad \gamma(n) = y(n) - y(n|n-1)$

ここで、$y(n|n-1)$は、$y(n)$の条件付平均であって、normal scalingを用いているのでガウス分布に従がい、次のように表わすことができる。

$$y(n|n-1) = E\{y(n) \mid Y(n-1)\}$$
$$= \langle y(n) Y^T(n-1) \rangle \langle Y(n-1) Y^T(n-1) \rangle^{-1} Y(n-1)$$

まず、条件付平均操作$E\{\cdot \mid Y(n-1)\}$を(5)の両辺に行えば、

(7) $\quad A_0 y(n|n-1) + \sum_{i=1}^{M} A_i y(n-i) = \sum_{i=0}^{M-1} B_i f(n-i|n-1)$

となる。残る手続きは$f(n-i|n-1)$を$\gamma(n)$で表現することであるが、これには$f(n)$の因果律と白色性を用い、行列計算を実行すれば次式が求まる。[7]

(8) $\quad f(k|m) = \sum_{i=0}^{m-k} R_k(m-i)\, \gamma(m-i)$

$$R_k(m-i) = V\, W^T(m-k-i)\, \Gamma^{-1}(m-i) - \langle f(k) Y^T(m-i) \rangle \Theta(m-i)$$

$$W(m) = H \Phi^m$$

$$\Gamma(m) = \langle \gamma(m) \gamma^T(m) \rangle$$

$$\Theta(m) = \langle Y(m-1) Y^T(m-1) \rangle^{-1} \langle Y(m-1) y^T(m) \rangle \Gamma^{-1}(m)$$

(5),(7),(8)を組合せると、次の時系列モデルを得る。

(9) $\quad A_0 y(n) + \sum_{i=1}^{M} A_i y(n-i) = A_0 \gamma(n) + \sum_{i=1}^{M-1} C_i \gamma(n-i)$

$$C_i = \sum_{\ell=i}^{M-1} B_\ell\, R_{n-\ell}(n-i)$$

このようにして、AR·MA(autoregressive moving average)型の

時系列モデルが求まり、モデルの時間おくれの次数M、各行列係数A_i、C_iも定まることになる。 又、係数内に含まれている$\Gamma(m)$の時間発展を(5)(9)を導置して、両辺平方し、fとrの直交性をそれぞれ利用すれば、次式(generalized Einstein relation)がまる。

(10) $$A_0 \Gamma(n) A_0^T + \sum_{i=1}^{M-1} C_i \Gamma(n-i) C_i^T = \sum_{i=0}^{M-1} B_i V B_i^T$$

§3. 情報の縮約

§2の計算手順は、基礎方程式(2)においてまず非可観測量を消去し、続いて、観測データ空間への射影を行って、時系列モデルを得たことになる。 有限次元の線形方程式では、この手順を逆にすることが可能となる。 まず(2)の観測データ空間への射影は、制御理論におけるカルマンフィルター[8]の導出過程(観測ノイズを含まない場合)と同じである。 つまり、(2)は次のようになる。

(11) $$x(n|n-1) = E\{x(n)\,|\,Y(n-1)\}$$
$$= \Phi x(n-1|n-2) + \Phi P(n-1) H^T \Gamma^{-1}(n-1)\, r(n-1)$$

よって、

(12) $$x(n|n) = \Phi x(n-1|n-1) + P(n) H^T \Gamma^{-1}(n)\, r(n)$$
$$P(n) = \langle \varepsilon(n)\, \varepsilon^T(n) \rangle$$
$$\varepsilon(n) = x(n) - x(n|n-1)$$

一方、観測の方程式も、同じ操作を通じて、

(13) $$y(n) = H\,x(n|n)$$

となる。次の操作は、非可観測量の消去である。(11)と(13)から、$x(n|n)$を消去すれば同じ(9)をもめることができる。この証明は省くが、([7]も参照)、その結果としては操作手順に関する可換図としてまとめることができる。

$$\begin{tikzcd}
\text{Langevin eq. }(2) \arrow[r, "P"] \arrow[d, "E"'] & \text{Kalman filter }(12) \arrow[d, "E"] \\
\text{non-Markovian eq. }(5) \arrow[r, "P"] & \text{AR·MA model }(9)
\end{tikzcd}$$

region I　　　　region II

P : projection

E : elimination

図、情報の縮約と可換図

つまり、図のregion IIにおいて、時系列データから何らかの情報量規準により、正確に時系列モデルが定められたとすると、(9)なるモデルを手に入れたことになる。矢印と逆の手順を行えば(12)にたどりつくことができるが、(2)まで遡ることはできない。つまり、非観測量を観測量の過去の情報で等価的にとりもどすことができたとしても、region

Iの元の方程式まではもどっていない。 これは、データをもとにしてregion IIでマルコフ性は回復されたとしても、region Iからregion IIへの射影操作が含まれているためである。 region Iでは、確率変数$x(n)$の動きを調べているが、region IIでは、条件付確率変数$x(n|n-1)$を調べていると言えよう。 そこで、この両者の情報量の差としては、次の相互情報量の形で表わすことが適当であろう。

(14)
$$
\begin{aligned}
I_n &= -\int P(x(n)) \log P(x(n))\, dx(n) \\
&\quad + \int P(x(n), Y(n-1)) \log P(x(n)|Y(n-1))\, dx(n)\, dY(n-1) \\
&= E\left\{ \log \frac{P(x(n)|Y(n-1))}{P(x(n))} \right\}
\end{aligned}
$$

最後に、この量Iは、正規性と定常性を用いて次式にも書き直しうる。

(15)
$$
I_\infty = \frac{1}{2} \log \det\left(\Sigma(\infty) P^{-1}(\infty)\right) \qquad \Sigma = \langle x x^T \rangle
$$

以上のことから、微視系の統計物理だけでなく、巨視系においても状態変数の縮約という考え方は、一つの重要な見方といえよう。 本稿が、非線形系、空間依存の問題のための一つの足場となれば幸いである。

文献

[1] H. Mori, Prog. Theor. Phys. 33, 423 (1965).

[2] H. Akaike, Ann. Inst. Statis. Math. 22, 203 (1970).
idem: IEEE. Trans. Automat. Control AC-19, 716 (1974).

[3] N. G. van Kampen, Can. J. Phys. 39, 551 (1961).
R. Kubo, K. Matsuo and K. Kitahara, J. Stat. Phys. 9 51 (1973).
K. Tomita and H. Tomita, Prog. Theor. Phys. 51, 1731 (1974).

[4] K. Kishida, S. Kanemoto and T. Sekiya, J. Nucl. Sci. Technol. 13, 708 (1976).

[5] K. Kishida and H. Sasakawa, J. Nucl. Sci. Technol. 17, 16 (1980).

[6] 古田勝久，線形システムの観測と同定　コロナ社.

[7] K. Kishida, 投稿中.

[8] R. E. Kalman, J. Basic Engr. (Trans. ASME, D) 82, 34 (1960).